# IO Intensive Applications at PNNL

Bruce Palmer

Annette Koontz

Karen Schuchardt

Pacific Northwest
NATIONAL LABORATORY

# IO Intensive Applications

▶ Development of an IO layer for a high resolution Global Cloud Resolving Model (GCRM)

▶ Development of a parallel IO layer for large simulations of a Smoothed Particle Hydrodynamics model for simulating flow in porous media
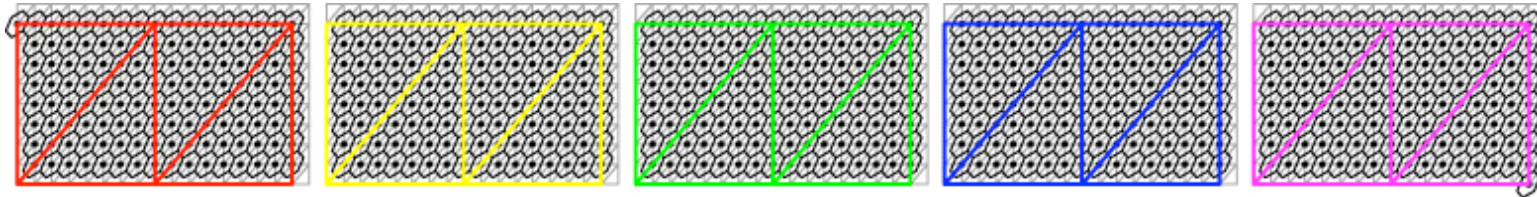
# Geodesic Grid
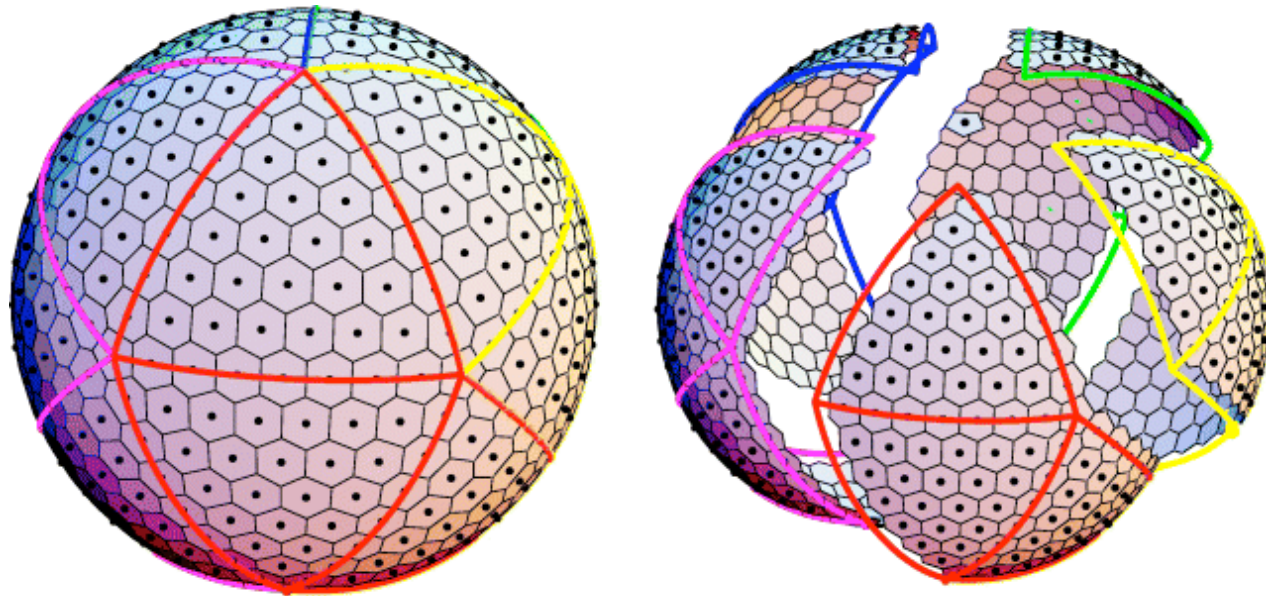


**Geodesic grids can be decomposed into 5 2x1 panels with the north and south poles left over**

# Geodesic Grid Panels
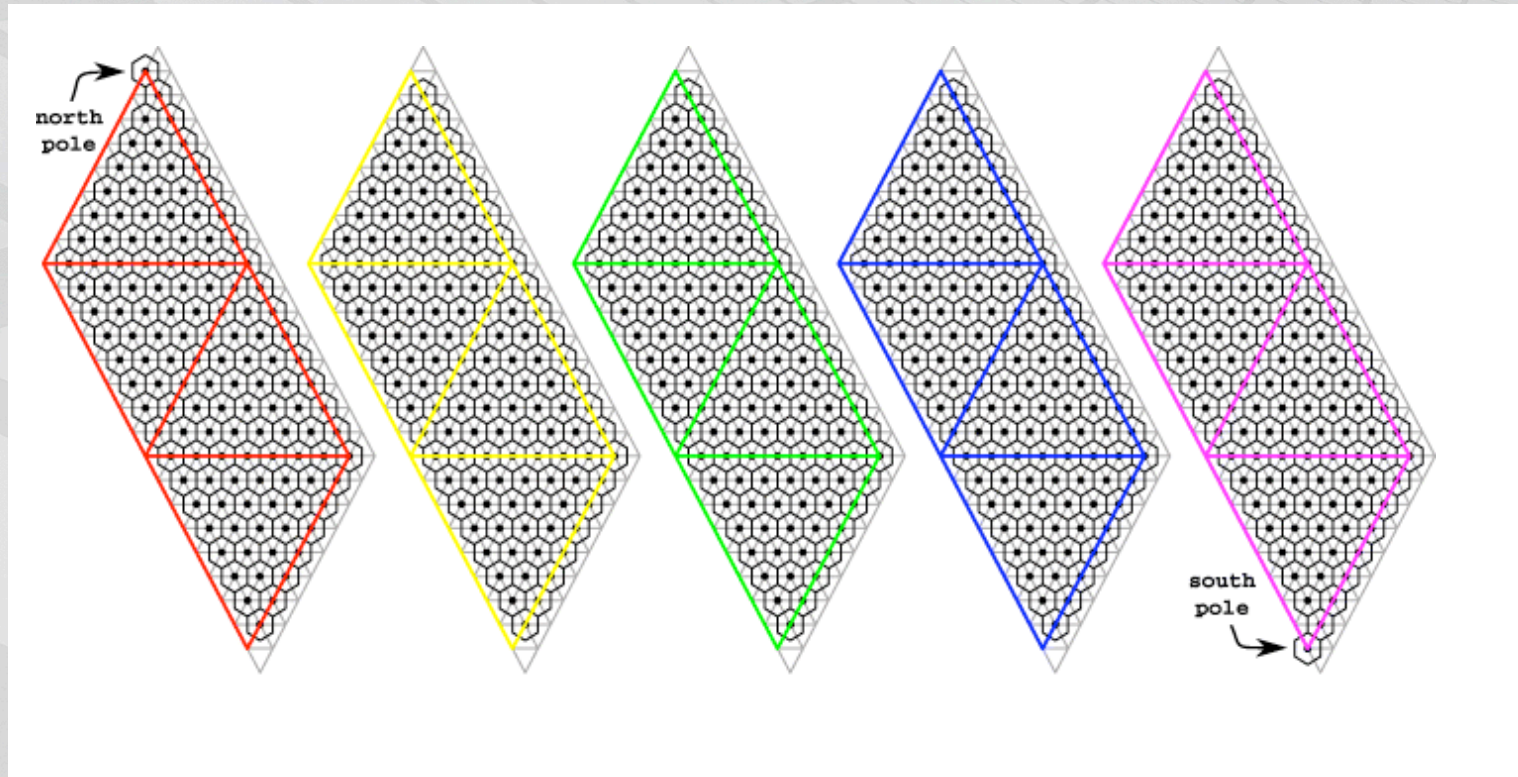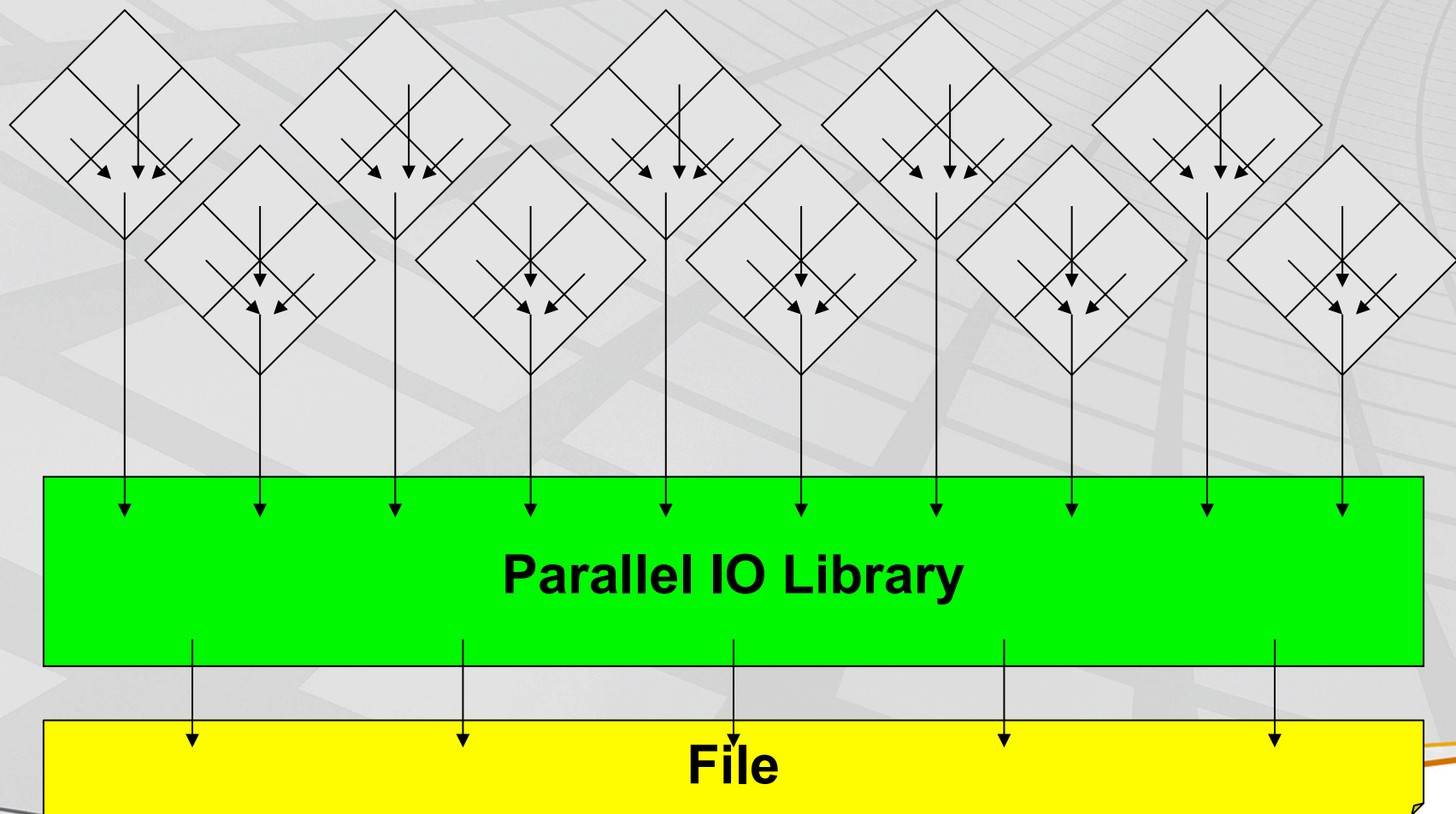
# Geodesic Grid Data Layout



5 2x1 panels of data or 10 square panels plus the north and south poles. The square panels are dimensioned by an integral power of 2. The total number of grid cells is given by

$$N = 10 * 2^{2R} + 2$$

where R is an exponent that characterizes the resolution of the grid. 4km resolution corresponds to a value of R = 11.
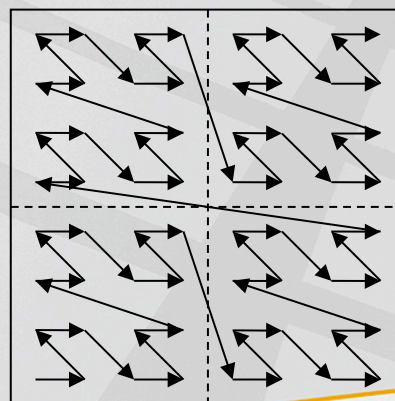
**Pacific Northwest**
NATIONAL LABORATORY

# IO Model for the GCRM



**Parallel IO Library**

**File**

Pacific Northwest
NATIONAL LABORATORY

# File Layout



Poles

Panels

N  S  0  1  2  3  4  5  6  7  8  9

Morton-ordering
within panels

Pacific Northwest
NATIONAL LABORATORY

# GCRM Targets

▶ Run GCRM at 4km resolution or less

▶ 16 GBytes of data per cell-centered variable per snapshot.

  ■ Corner centered data is an extra factor of 2 larger, edge centered data is 3 times as big

▶ Produce snapshots every simulated hour

▶ Write out ~20 variables per snapshot

▶ Simulate model for 1 year

▶ Produce ~1-10 Petabytes of simulation data

**Pacific Northwest**
NATIONAL LABORATORY

# IO Targets for GCRM

▶ Add less than 10% overhead to cost of running GCRM model

▶ Produce files that are easy to access for subsequent analysis and processing (i.e. files that are in standard formats etc.)
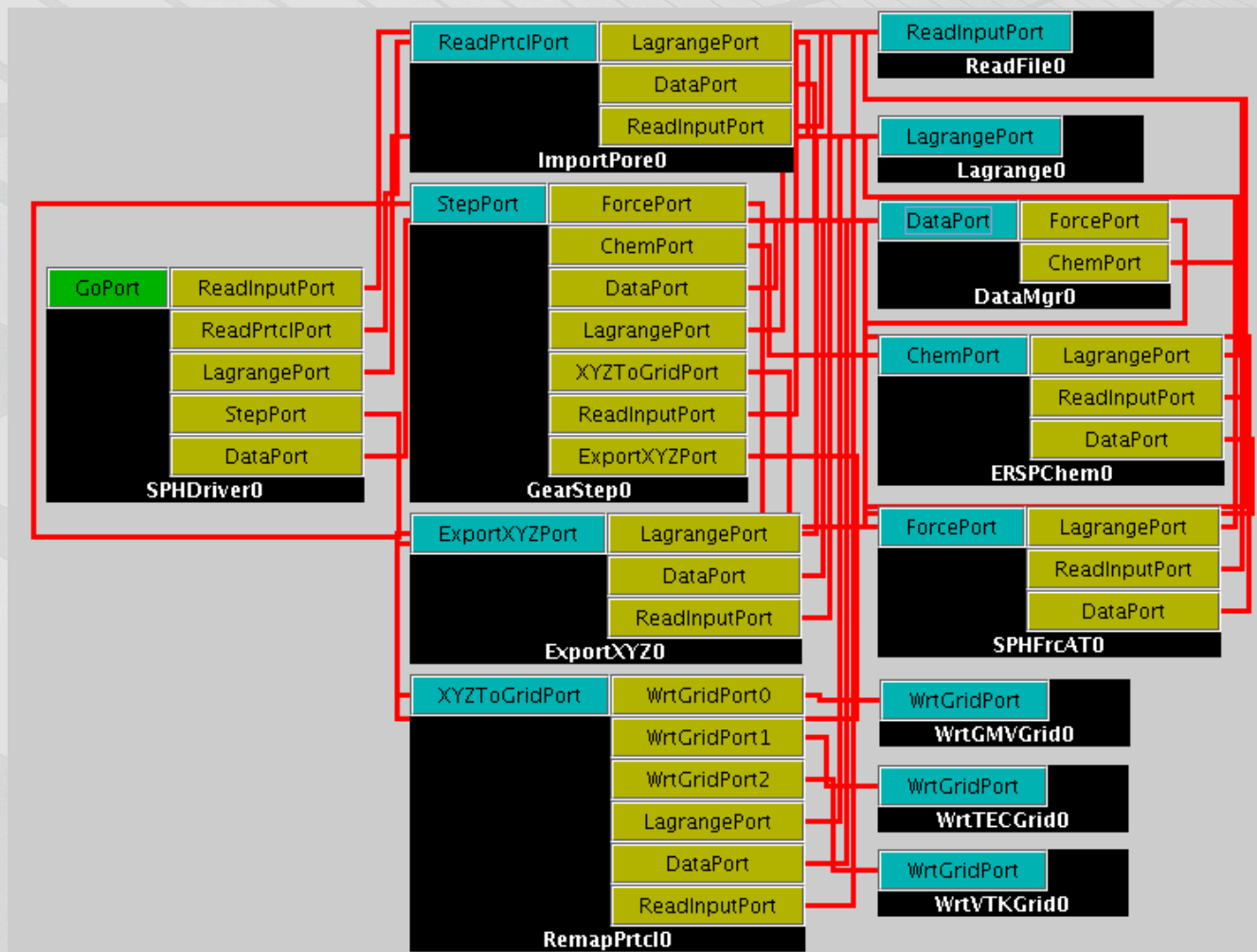
**Pacific Northwest**
NATIONAL LABORATORY

# IO Targets for GCRM (cont.)

| R | Cell width (km) | Number of Cells | Number of Corners | Number of Edges |
|---|---|---|---|---|
| 10 | 8 | 10485762 | 20971524 | 31457266 |
| 11 | 4 | 41943042 | 83886084 | 125829126 |
| 12 | 2 | 167772162 | 335544324 | 503316486 |
| 13 | 1 | 671088642 | 1342177284 | 2013265926 |

# SPH Groundwater Framework

▶ Developing Smoothed Particle Hydrodynamics and Continuum Darcy Flow Components for simulating multiscale reactive groundwater flows

▶ Framework uses the Common Component Architecture (CCA) to connect components together

▶ Need to output large lists of particles and their attributes as well as grid-based visualizations of data

**Pacific Northwest**
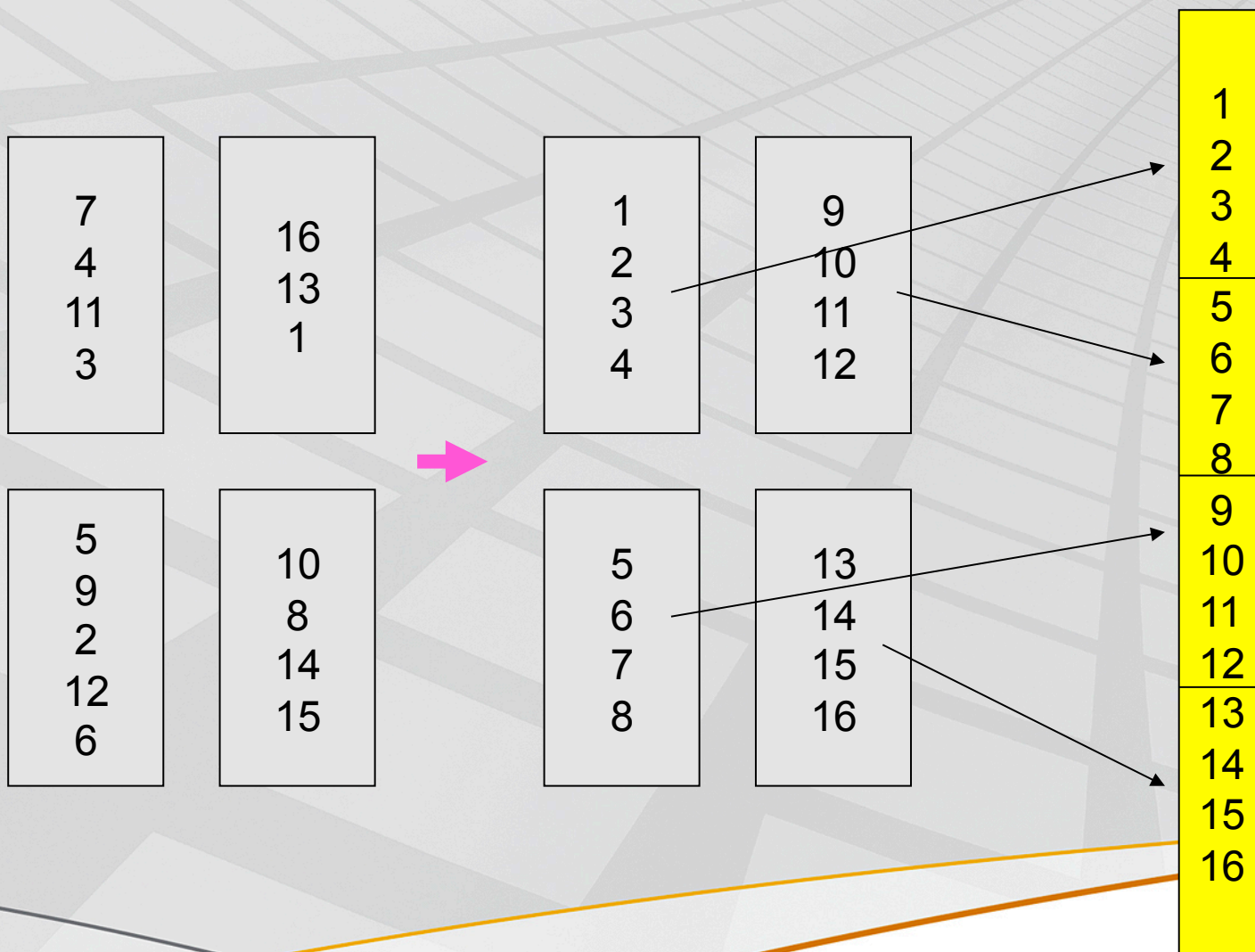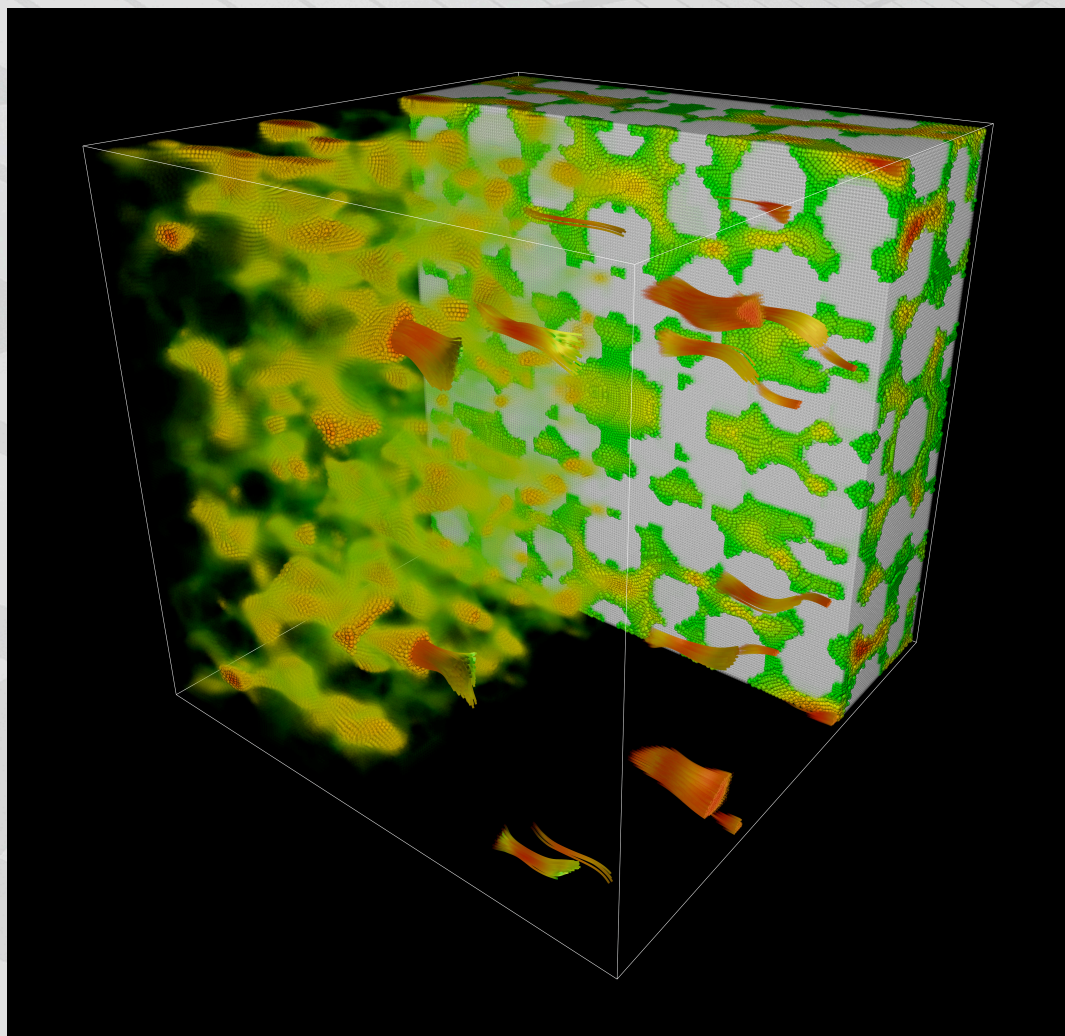NATIONAL LABORATORY

# SPH Framework

# Particle List Output
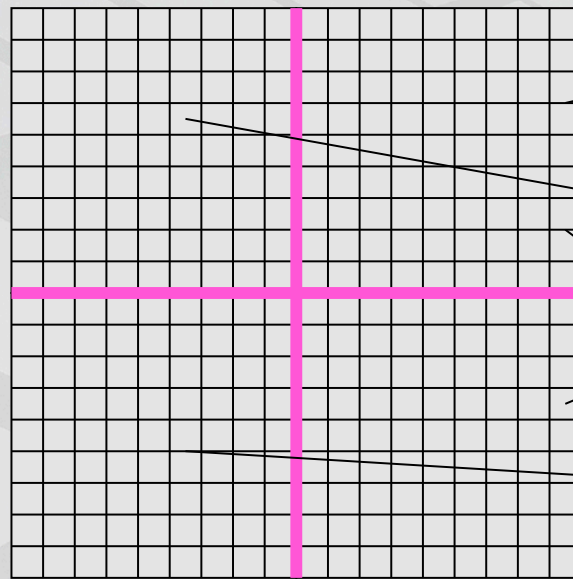
Type  x  y  z  attr1  attr2  attr3  .....

# Particle List IO

# Particle Output
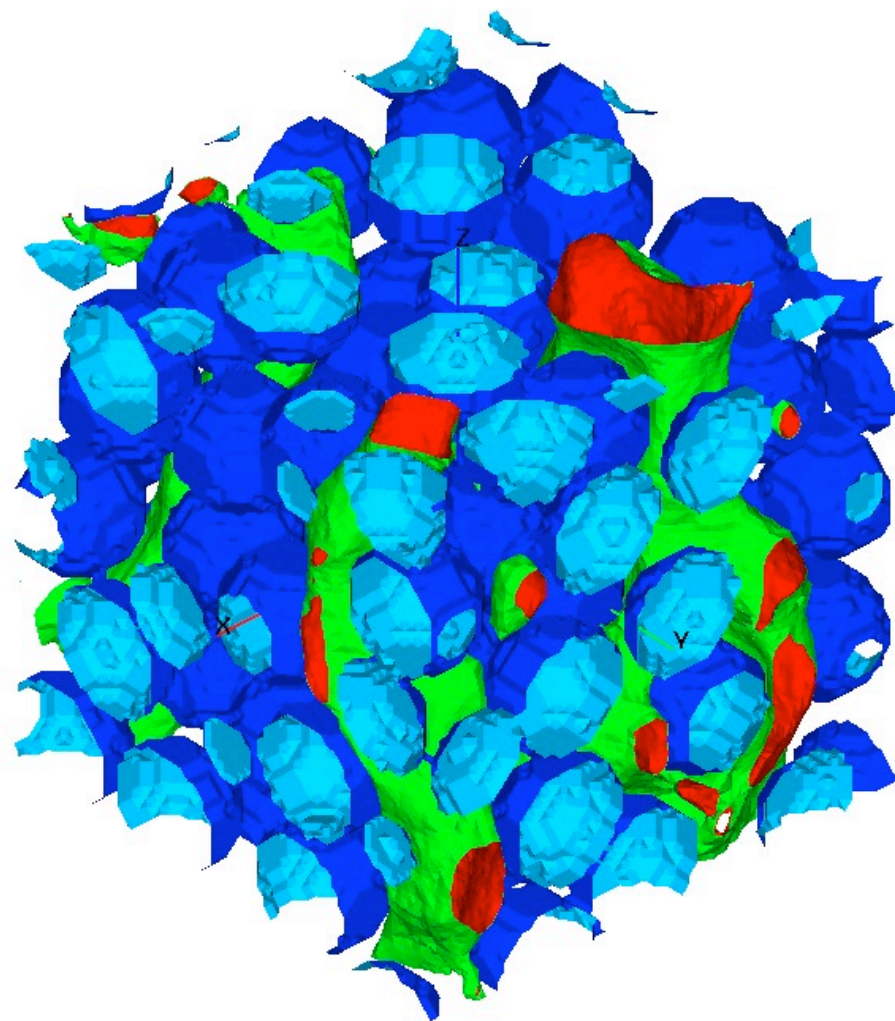
# Grid-Based Output



Smoothed Data

File

Pacific Northwest
NATIONAL LABORATORY

# Grid Output

# Continuum Darcy Flow Simulations

▶ Simulations on unstructured grids

▶ Cell-centered data in output

Pacific Northwest
NATIONAL LABORATORY

# Groundwater Simulation Targets

► Currently simulating ~7 million particles on 512 processors. IO is not a big issue yet but would like to get to simulations in the 100-1000 billion particle range or use more processors to simulate longer periods

► Targeting continuum simulations on the order of a billion grid cells

**Pacific Northwest**
NATIONAL LABORATORY